DOCUMENT RESUME

ED 069 728 TM 002 180

AUTHOR Helmstadter, G. C.
TITLE A Comparison of Traditional Item Analysis Selection Procedures with Those Recommended for Tests Designed to Measure Achievement Following Performance Oriented Instruction.
PUB DATE Sep 72
NOTE 16p.; Paper presented at American Psychological Association Convention (Honolulu, Hawaii, September 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Tests; Comparative Analysis; *Criterion Referenced Tests; *Discriminant Analysis; Instructional Design; *Item Analysis; Measurement Techniques; Performance Criteria; Performance Factors; *Test Construction

ABSTRACT

Several alternative indexes of item usefulness were empirically compared. They were: (1) item discrimination based on high and low groups on a postinstructional measure, (2) shift in item difficulty from a pre- to a postinstruction measure, and (3) item discrimination based on pre- and posttest performances. A typical classroom final examination was administered to 28 students on their first day in a multivariate statistics course and again at the end of the term. The results, using the various indices, indicated that shift in item difficulty from pre- to postinstruction yielded data significantly more similar to the pre-post discrimination index than did the high-low group posttest discrimination index. It is recommended that the conceptually more ideal pre- to postinstruction discrimination index be used even when instruction is not performance oriented. (Author/JS)

A Comparison of Traditional Item Analysis Selection Procedures with Those Recommended for Tests Designed to Measure Achievement Following Performance Oriented Instruction

G.C. Helmstadter, Arizona State University, Tempe, Arizona

Several alternative indexes of item usefulness were empirically compared: 1) item discrimination based on high and low groups on a post instructional measure; 2) shift in item difficulty from a pre to a post instruction measure; and 3) item discrimination based on pre and post test performances.

Shift in item difficulty from pre to post instruction produced results significantly more similar to the pre-post discrimination index than did the high-low group post test discrimination index. It was recommended that the conceptually more ideal pre to post instruction discrimination index be used even when instruction is not performance oriented.

A Comparison of Traditional Item Analysis Selection Procedures with Those Recommended for Tests Designed to Measure Achievement Following Performance Oriented Instruction

G.C. Helmstadter, Arizona State University, Tempe, Arizona

Introduction

Popham and Husek (1969) have stated that traditional item analysis procedures may not be useful for constructing achievement tests in situations where the instruction has been performance oriented. Rahmlow, Matthews and Jung (1970) have further argued that in such criterion referenced situations the homogeneity of post instruction performance may be so great when the instruction has been successful that discrimination indexes may not be useful in the traditional sense. Rather, these latter authors suggest that the function of a discrimination index in a criterion referenced situation is primarily that of indicating the homogeneity of the item with respect to the specific instructional objective measured. Further, as an alternative to the traditional interpretation of a discrimination index, these latter authors focus attention on a shift in item difficulty from a non-instruction to a post-instruction measure.

Unfortunately, the arguments offered against the use of traditional procedures have assumed that discrimination indexes must be based on a comparison of high and low performers on a post instructional measure and generally ignored the possibility that such an index might be based on a

difference in response to the item from a pre instruction to a post instruction measurement. Thus it is entirely possible that the alternative procedure often proposed by advocates of criterion referenced measurement of using the change in item difficulty from pre to post instruction measurement may be less efficacious in selecting items for achievement measures than the traditional item discrimination index applied in the more ideal way of contrasting responses before and after learning.

The purpose of this study was, therefore, to empirically compare the following three alternative indexes of item usefulness: 1) the item discrimination based on high and low groups on a post instructional measure; 2) the shift in proportion of subjects getting the item correct from a pre to a post instructional measure; and 3) the item discrimination based on pre and post test performances.

Method

The procedure followed was to administer a typical classroom final examination the first day of class to 28 students enrolled in a multivariate statistics course and again to the same subjects at the end of the term. The traditional item indexes of difficulty and of discrimination were computed in the usual way on both the pre instructional and the post instructional responses to the item. An additional item discrimination index was then computed for each item comparing post instructional responses with pre instructional responses. Finally, the change in item difficulty from pre to post instruction administrations was determined for each item.

The extent to which the three different recommended item selection procedures produced similar results was determined by computing the intercorrelations among the three indexes over the 59 items used in the examination.

Using a selection criterion of .20 and up for the discrimination indexes and of .10 or more increase in the proportion who got the item right on the post test as compared with the pre test, cross-tabulations were made to determine how many of the items selected and rejected by the more ideal procedure were similarly selected and rejected by the traditional high-low group post test discrimination index and by the shift in proportion correct index advocated by many persons using criterion referenced measures.

Results

A statistically significant ($P<.001$) increase in mean test scores for pre to post instruction of 19.2 and a change in average item difficulty from .28 to .59 indicated that changes in performance on the measure used did indeed occur.

The correlation between the two discrimination indexes was found to be .47; that between the pre to post discrimination index and the shift in proportion correct was .78; and that between the post test high-low discrimination index and the shift in difficulty was .29. The difference between the correlations of .78 and .47 as determined by a formula appropriate for situations where the correlations are both calculated on the

same subjects was statistically significant ($.001<P<.005$).

The cross tabulations indicated that of the 59 items included in the examination, 67 percent would have been similarly selected and rejected by all three indexes; 71 percent would have been similarly selected or rejected by the pre to post discrimination index and the high-low discrimination index; and 90 percent would have been similarly selected or rejected by the pre to post discrimination index and the shift in proportion correct index.

Discussion

These data clearly confirm the contention of those who have urged caution in using traditional item analysis procedures in criterion referenced situations. The results further suggest that even when the instruction is not performance oriented the use of a high-low group discrimination index obtained on a post instruction measure may not provide an appropriate index for the selection of items to be used in constructing an achievement test.

While the use of the change in item difficulty from pre to post instruction can lend to the selection of nearly the same set of items as the use of a pre to post instruction discrimination index, there seems to be no advantage to the use of shift in proportion who get the item correct over the more conceptually satisfying pre to post instruction discrimination index since both require two separate administrations of the pool of items from which the selection is to be made.

References

Popham, W.J. & Husek, T.R. Implications of criterion-referenced measures. Journal of Educational Measurement, 1969, 6, 1-9.

Rahmlow, H.F., Matthews, J.J., & Jung, S.M. An empirical investigation of item analysis in criterion-referenced tests. A paper presented at the annual meeting of the American Educational Research Association, Minneapolis, Minnesota, March 6, 1970.